



# Counternarrative Campaign for Preventing Radicalisation

# Deliverable D5.2 Campaign effectiveness evaluation report 1

**Authors:** 

Ronald Denaux; Angelo Rollo

> This project was funded by the European Union's Internal Security Fund —Police under GA N° 812613







- Università degli Studi di Torino (UNITO) Project LEAD Coordinator;
- Society Against Violent Extremism (SAVE);
- Centro Studi Internazionali (CeSI);
- Zanasi & Partners (Z&P);
- Novareckon (NR);
- Tecoms Srl (TEC);
- European Foundation for Democracy (EFD);
- Confederazione Islamica Italiana (CII);
- Expert System Iberia (ES);
- Inoftron Europa SL (IE)



# Version history

Version	Date	Author	Description
0.1	05/07/2019	R. Denaux	Table of Contents
0.2	26/07/2019	R. Denaux	First draft
0.3	28/07/2019	R. Denaux	Final revision for submission
0.4	31/07/2019	A. Gunjak, P. Colla	Quality check before submission



# Table of Contents

1.	IN	ITRODUCTION	5
	1.1	Relation to Other WPs and Tasks	5
	1.2	Objectives	6
2.	T	OOLS AND INSTRUMENTS	7
	2.1	Social media metrics	7
	2.2	Content collector and text analysis	8
3.	B	ASELINE EVALUATION	24
	3.1	Monitored Sources and Existing Campaigns	24
	3.2	Social media analytics	25
	3.2	Text analytics	26
4.	C	ONCLUSIONS AND FUTURE WORK	32



# 1. INTRODUCTION

The main output of the Cicero Project will be the Cicero counter-radicalisation campaign. This document is one of the first WP5 deliverables. WP5 deals with campaign evaluation and this document is the first of three deliverables planned within Cicero to document the effectiveness monitoring and evaluation of the campaign. However, since the campaign has not started yet, in this document we illustrate some of the evaluation tools that will be used.

# 1.1 Relation to Other WPs and Tasks

The main inputs for this deliverable are:

- D5.1 Qualitative and quantitative evaluation methodology, which is the result of work performed as part of T5.1. In essence, T5.2 is an implementation of this methodology. Note that by the time this document is being written, the methodology is still being defined. Therefore, this document only tests some of the tools identified in the evaluation methodology.
- Another important input for this work is WP3, which is responsible for defining the counternarrative strategy and for generating the content that should be included as part of the campaign. Knowing the strategy and contents that will be produced is crucial for being able to know what should be monitored and to identify what aspects (e.g. topics, sentiments) will be good indicators of the campaign effectiveness. As with the evaluation methodology, the D3.1 CICERO counternarrative strategy is still being defined at the time this document is being written, therefore, in this iteration of the document, we will only be able to show tool capabilities based on pre-existing campaigns.
- WP4 is responsible for the campaign dissemination; the initial dissemination plan was documented in D4.1 "Dissemination Plan No 1" which provided a high-level overview of the tasks needed for both project and campaign dissemination. Regarding campaign-specific dissemination, which is the more relevant part for this document, D4.1 identified a high-level (social-)media strategy as well as reliance on a committee of young people who will play a crucial role in dissemination of the campaign. At the time of writing a second version of the dissemination plan, D4.2, is being written. However, because the identification of target audience has not been finalized yet and the campaign strategy is still in development, no specific sites, topics or hashtags are currently available. This means that, in this evaluation report, we are as yet unable to start monitoring content that will be representative of the CICERO campaign.
- Although WP2 does not provide direct input to this task, the analysis of the various types of extremism described in D2.1 heavily influences the counter-narrative strategy and thus also affects how we will evaluate the effectiveness of the campaign. In particular, the prevalent narratives and topics identified in D2.1. Also, the network of experts identified and documented



in D2.2 will play a crucial role in advising and shaping the campaign's objectives and hence we will take that advice into account when deciding which metrics should be monitored and which methodology should best be used.

# 1.2 Objectives

Since the Cicero campaign has not started yet, in this document we provide preliminary sample results using existing campaigns and tools that will be used to monitor the Cicero campaign once it starts. The main objectives are to:

- Choose tools and validate their capabilities and limitations with regard to the evaluation methodology
- Obtain a baseline evaluation prior to the start of the Cicero campaign. This can serve to put measurements obtained after the Cicero campaign begins into context.



# 2. TOOLS AND INSTRUMENTS

D5.1 describes the overall methodology that will be used to assess the effectiveness of the campaign. The methodology relies on various tools and instruments to collect and measure key qualitative and quantitative metrics that are indicative of the success of the campaign. Since, at the time of writing, the campaign and dissemination strategy, are still ongoing, in this section we describe a couple of tools we will configure and use during the evaluation phase. Note that once more details about the campaign and dissemination strategies are decided, we may decide to include other tools if we deem the current tools to be insufficient. However, the tools presented below are configurable enough that we are confident that they will provide valuable metrics for the campaign evaluation.

#### 2.1 Social media metrics

The impact of CICERO on its intended audiences will be measured according to two main factors, which can be assessed and quantified; these two factors will be awareness and engagement.

These factors will be used jointly, in order to delineate the overall impact of the activities of the CICERO campaign.

#### / Awareness

The awareness metrics can calculate how many viewers are reached by the campaign content online, and their characteristics. Furthermore, these metrics can aid in the definition of the modalities, the time and the location in which the audience has been exposed to the messages of CICERO campaign, and can highlight which part of the campaign has mostly affected them.

As explained within D5.1, there are several metrics that are used to calculate the awareness raised by the campaign; these are: Impression, Reach, Impression frequency and Views.

#### // Engagement

Engagement metrics differ from awareness metrics in that whereas the latter can estimate the viewers reached by the campaign, the former can determine how the campaign message has been received, and the consequent viewer's reaction, thus delivering precious information about the impact of CICERO campaign.

Engagement metrics will be also exploited to detect which part of the campaign has obtained the largest impact.

As shown by D5.1, the engagement metrics are: Audience retention, Likes, Comments, Bounce, and exit rates, Shares and Sustained engagements.



# 2.2 Content collector and text analysis

Within the Cicero project, Expert System Iberia (ESI) has configured an instance of its commercial product, the Analysts' Workspace (AW) to be used for monitoring both radical and counterradical content published on- and off-line. In this section, we briefly explain what the Analysts' Workspace is and what can be done with it, in particular when applied to monitoring of (counter-) radical content.

In a nutshell, the AW makes it possible to:

- monitor and collect on-line and off-line textual content from a variety of sources. Examples of on-line sources: website-feeds for changes to websites, social-media APIs for social-media content, webcrawls for pre-existing websites, Google alerts and Queryon-the-Fly for new web-content that matches certain keywords. Off-line sources include PDF and Word documents that can be uploaded to the system.
- Automatically analyse the collected textual content. As a result, it will become easier to find and group analysed content based on various topics and entities.
- **Manually analyse the collected content** in order to discover trends and answer specific questions over a large number of documents.
- Write analysis reports and disseminate them as part of a team.

In the next subsections, we'll go into more details about each of these capabilities; most of the described details may be too technical for the non-technical reader or for them who are not going to be using the system to perform analysis. For non-technical readers, we suggest to skip the subsection I to V and continue reading from subsection VI.

# /// Architectural Overview

**Errore.** L'origine riferimento non è stata trovata. shows the overall architecture of the AW as adapted for C ICERO. In this section, we shortly describe each of the components, but most of them are described in more detail in subsequent sections.

As the figure shows, the AW is built on top of a **distributed configuration** layer, implemented using Apache Zookeeper. This layer ensures that the overall service is scalable because all the components can be configured and deployed in a distributed manner; this is important for ensuring that the application will be able to function regardless of whether we need to monitor only a few dozen documents per day or we need to monitor tens of thousands of documents per day. Being distributed makes it possible to deploy the whole system on a single server, but we can also choose to use a cloud-based deployment where there are several servers running each component.





#### Figure 1 AW for Cicero architecture overview

The core of the AW is the **Data Processing Engine**, which maintains a set of *sources* and target *document collections*. The data processing engine acts as the orchestrator and triggers new web crawls based on a schedule configuration and monitors when new crawled documents are added to their respective collections in the document index. When needed it triggers automatic machine translation of the raw texts and triggers semantic enrichment of the texts. As an extension to AW, as part of CICERO, we have added a custom analyser for the texts based on the stylometric analysis of the text. As a result, we can produce estimates about the education level and age of the authors of documents being collected.

The **Web Crawler** is responsible for crawling websites, retrieving the HTML content and filtering the raw textual content that can be analysed. The crawler also is responsible for extracting metadata that is related to the crawled website. The Web Crawler is implemented using Apache Nutch.

The **Document Index** is software that is able to create and manage various document databases, called collections. These collections are stored in such a way that they can be searched quickly. We use Apache Solr as the underlying implementation for the document index. Within CICERO we currently maintain 5 collections, one for each of the types of monitored extremism (salafi-jihadism, far-right, far-left, eco-terrorism) and one for monitoring pre-existing counter-radicalisation campaigns. Each of these collections is populated through a list of relevant sources (the various implemented source types are described below) for the specific type of extremism or based on keywords related to existing campaigns. Besides



these 5 main CICERO collections, we expect that we will need to create a new custom collection for monitoring the CICERO campaign, based on hash-tags, topics, and target audiences once they are decided as part of the campaign and dissemination strategies.

The **Semantic API** is responsible for performing semantic enrichment of the text contents. This component is based on Expert System's Cogito technology and described below in further detail.

Finally, the overall AW service is accessible via a powerful, yet easy to learn, web interface. This means that CICERO partners can log into the application via their browser and add new sources, search and explore the collected documents and find metrics relevant for determining sentiments, topics and narratives that are being used in radical and counter-radical sites. that programmers can access the service via commonly used web protocols (HTTP and HTTPS) and process the available information using the JSON data exchange format. This again is very common and makes it easy to search the collected content. The JSON format also makes it very easy to build web and mobile application interfaces on top of this service that end users can utilize. Section 4 describes this interface in detail.

# *IV* Content source types

AW can collect content from a variety of sources based on a variety of technical protocols. More specifically, we support RSS feeds, Twitter sources, keyword-based sources and manual uploading of documents. We describe these in more detail below.

For each of the source types, the content collector can be configured to specify how often the source should be crawled. By default, we crawl sources once a week but depending on the source, we can choose to crawl every hour (e.g. when following a trending topic on Twitter), or only once (e.g. when crawling an existing website that is no longer being updated).

RSS feeds: We support defining sources based on one or more RSS feed URLs. RSS is a web protocol that allows you to "subscribe" to updates on a website. I.e. when a site adds new content, the feed is updated, allowing the AW to fetch and analyse the new content. Each source can be given a unique name, that will be shown as part of the collected content metadata. For example, we can define a "Salafi Jihadism RSS" source, which points to various RSS feed URLs for sites that have been identified by radicalisation experts (e.g. khilafah.com). This allows for fine-grained definition of sources. RSS can also be used in combination with Google Alerts<sup>1</sup>, which allow you to define search terms (possibly in combination with filters for a specific language and region). Google then publishes new websites that match those keywords as an RSS feed, which can be ingested by AW (this is effectively similar to a keyword-based source, described below).

<sup>&</sup>lt;sup>1</sup><u>https://www.google.com/alerts</u>



- Twitter API Sources: AW supports defining sources based on the Twitter developer API. We support the basic standard (i.e. non-premium) operators<sup>2</sup>, which include defining source per keywords, phrase, disjunction (OR), exclusion, hashtag, account, dates, replies to an account, Twitter-provided filters (e.g. potentially sensitive content, image or video). As with RSS feeds, we can associate multiple Twitter keywords to a single Twitter source for added flexibility (e.g. to search for keywords in different languages), or to modify the queries without needing to define new sources.
- Facebook API Sources: AW supports defining sources based on a Facebook app that is created through a Facebook Developer Account. This app can be used to crawl posts and comments from public pages and groups. Although technically possible, since the Cambridge Analytica scandal, it is almost impossible to get Facebook's permission to connect to their API; therefore it is unlikely that we will use this type of sources.
- Keyword-based sources: AW supports defining sources based on keywords, or to be more precise: on *search engine queries*. We have implemented support for the main search engines: Google, Bing and/or Yahoo. By default, every time this source is crawled, we keep the first 50 results returned by the selected search engine(s). Because we work at the level of queries, we can use the advanced query operators provided by the different search engines to define fined-grained sources. For example, we can use the site operator to only return results for a specific website (presumably one that does not support RSS feeds, or only offers a very broad feed), match exact phrases using double quotes and exclude words using the minus(-) character. Each search engine has its own set of advanced search operators, thus we refer to the documentation of Google<sup>3</sup>, Bing<sup>4</sup> and Yahoo<sup>5</sup>.
- Nutch WebCrawl sources: Since AW's implementation is based on Apache Nutch, it is also possible to create a source based on a Web Crawl. In this case, the source is defined by one or more "seed" URLs which are used as starting points. The crawler then adds those websites, but also collects URLs which are linked from those seed pages and collects those as well in the following iteration. Therefore, besides the seed URLs, this type of sources also require the definition of a maximum iteration. Due to the exponential nature of this type of source, we typically recommend a low value for the maximum iteration (e.g. 2).

# V Machine Translation

Monitoring of content relevant for CICERO, both automatically and manually, is difficult due to the multiple languages that are being considered within the project (English, Italian, French and Spanish). In order to support the required languages, the AW has been configured to use a third-party machine

<sup>&</sup>lt;sup>2</sup> https://developer.twitter.com/en/docs/tweets/rules-and-filtering/overview/standard-operators

<sup>&</sup>lt;sup>3</sup> https://support.google.com/websearch/answer/2466433?hl=en

<sup>&</sup>lt;sup>4</sup> https://docs.microsoft.com/en-us/previous-versions/bing/search/ff795620(v=msdn.10)

<sup>&</sup>lt;sup>5</sup> https://search.yahoo.com//web/advanced



translation service provided by Systran<sup>6</sup>. In this section, we motivate the need for using machine translation and discuss the advantages and disadvantages of this approach.

In short, the main motivation for using ML is that providing native semantic enrichment for each individual language is prohibitively expensive. To understand why we need to explain how the Semantic Enrichment performed by AW works. The semantic enrichment is based on Expert System's Cogito technology. Although Cogito provides native support for 14 languages, not all of these 14 languages have the same level of support. At its core, Cogito performs the usual NLP tasks of tokenization, part-of-speech detection, lemmatization, and sentence splitting. Then Cogito performs word sense disambiguation. After that, rules are executed to extract (named) entities and to perform categorization (the results of which are what is typically understood as enrichment). The main issue is that "native support" is considered achieved when Cogito is able to perform word sense disambiguation; however the most mature languages (English, Italian, Spanish) each have over 100K manually crafted rules to perform NER and Categorization for our main taxonomies (Intelligence, Crime, Cybercrime, Terrorism, Geography and Emotions), which are needed to achieve a high level of precision and recall. Developing such rules for other languages requires person-years and is an effort that is out of the scope of Co-Inform.

By using ML in this project, we can use English as our core language for Semantic Enrichment, and we will use a Machine Translation service to translate from Italian, French, Spanish (and a few other languages) to English. This is a common solution that works well in practice. Although the results are not as good as they would be if we had mature native support for these languages, the quality of machine translation is currently good enough to be useful. In general, while accuracy for English texts is around 90%, accuracy for translated texts can be in the range of 80 to 85%, depending on the quality of the translation and the quality of the text itself. Hence, this solution will be sufficient. Switching to a full native support for different languages can be done after the project as part of the exploitation of the project results if the extra 5 to 10% inaccuracy is desired.

One advantage of using a core language is that all texts can be inspected by English speakers, even if they do not speak the original language of the content (Italian, French, Spanish and even Arabic), making it easier to verify that technical services are working as intended.

# VI Semantic Enrichment via Cogito

Expert System's semantic text analysis technology is called Cogito. At its core, Cogito performs the usual NLP tasks of tokenization, part-of-speech detection, lemmatization, and sentence splitting. Then Cogito performs word sense disambiguation. After that, rules are executed to extract (named) entities and to perform categorization.

<sup>&</sup>lt;sup>6</sup> http://www.systransoft.com/



- Standard AW Categories: AW provides 6 standard taxonomies, each one defines a long list of categories or topics:
  - Intelligence: this is a generic taxonomy that contains categories relevant to the intelligence and security domain. Since many areas of disinformation are related to these categories, being able to filter by these will be useful. This taxonomy is organized in up to 4 levels of subcategories. Some of the broad categories are "Arts, Culture and Entertainment" and "Environmental Issue", while some of the more narrow categories include "Air Pollution", "Genetics" and "Housing and Urban Planning".
  - Crime: this is a taxonomy focused on crimes. This includes broad categories such as "Property crime" and "Positive results from investigations and measures"; and narrow categories such as "Immigration-related offenses", "Acquittals"
  - Cybercrime: this is a taxonomy focused on illegal activities online with broad categories such as "Cyber Attacks" and "Cyber Security" to narrow categories such as "Zero-day", "Ransomware" and "Defacing".
  - Terrorism: is a taxonomy focused on issues related to terrorism. In particular, it provides different ways to categorizing content via broad categories like "Terrorism by matrix", which has subcategories "Religiously inspired terrorism" and "Narco-terrorism". Other broad categories include "Terrorist activities and tactics" and "Counterterrorism".
  - Emotions: this is a relatively flat taxonomy (i.e. it is not nested into many layers of subcategories) that describes many emotional states such as "Joy", "Hope", "Forgiveness", "Confusion", "Regret", "Repulsion" and "Materialism".
  - **Geo**: provides categories for each country (and for the US for each state)
- Custom Categories: Not included in AW, but developed as part of other existing projects, we also include 2 custom taxonomies:
  - Strategic Radicalisation Narratives: this taxonomy was developed as part of the TRIVALENT EU-project to find radicalization narratives used by radical Islamist groups like ISIS and Al-Qaeda; however, the taxonomy is meant to be extensible to other types of extremism, hence we aim to extend this taxonomy as part of CICERO.
  - MediaTopics: this is a mid-sized taxonomy that focuses on generic topics related to news.
     This is included as part of CICERO to give a more generic set of topics since the standard AW taxonomies are mostly focused on intelligence and security topics.
- (Named) Entities: For (named) entity extraction, Cogito provides the standard named entities: People, Organizations, and Places. However, Cogito leverages additional information about known entities to provide more specific types with the additional advantage that in this case we can normalize the names. For example, Facebook can appear as "Facebook Inc", or simply as "Facebook".



### VII Socio-demographic analysis based on stylometry and language detection

Besides the semantic analysis performed by Cogito, as part of Cicero we are also performing additional analysis on the text. This is based on a combination of an existing Expert System module for stylometric analysis and custom machine learning models for age prediction and education level prediction. Another existing module as part of AW is a language detection module, which allows us to identify the language that the content is written in.

#### Stylometry for inferring education level

When analysing a text, Cogito is able to extract around 117 stylometric features from the text, these are measures of e.g. percentage of academic words, use of specific verb tenses, adjectives and adverbs. Based on this, we have trained a model that can provide a prediction about the grade level of the text's author, ranging from a numerical value of 2 (primary school) to 15 (post-graduate level). Below, we group these into four categories: primary, middle school, high school and college/university. This technology was developed prior to Cicero and is in production, suggesting that the results have an accuracy of around 80%.

#### Stylometry for inferring age-groups

We can use the same stylometric features to train a model to predict the author's age range. We did not have a previous implementation for this functionality, therefore this was developed as part of the Cicero project. We used an existing dataset<sup>7</sup> consisting of 681K blog posts for which the age of the author was known.

As any machine learning model, the results are not perfect and the results can contain errors. To get a feeling of how accurate the trained model is, we refer to the following histogram depicting the age difference between the predicted and the actual age of the author:

<sup>&</sup>lt;sup>7</sup> https://www.kaggle.com/tomlisankie/blog-posts-labeled-with-age-and-gender/version/1





2019

On the x axis we see the difference between the predicted and the actual age of the author and on the y axis we see the number of cases when this occurred. These numbers come from predicting the age for 132,000 blog posts. In 52% of the cases, the model predicted age within 6 years of the actual value. In about 20% of the cases, the difference was more than 15 years. In other words, the results of this model should be taken as indicative, but should not be used without looking at further evidence; e.g. by manually inspecting a sample of the documents.

If required we can train another model to predict the author's gender as well. Note that due to GDPR restrictions Expert System does not link these predictions to individual authors and CICERO partners are not able to see predictions for individual texts retrieved from the internet. However, we are able to provide aggregated predictions for (large enough parts of) collections, as we show below in Section 3.2.

#### VIII Content Storage and Indexing

After crawling, the extracted text and metadata are stored in a Solr database, more specifically in a collection in Solr. In parallel, the Data Processing Engine triggers a possible machine translation as well as the semantic analysis. When the semantic enrichment process is finished, the document in the Solr is updated by adding the semantic enrichment fields.

The Solr instance is configured in such a way that it is possible to query for documents based on keywords in the textual fields or by looking up substrings in the various semantic enrichments.

We also exploit the Solr functionality of **facet fields**, which provides a way to provide summaries of how many documents in the search result have specific values for specific fields. For example, suppose we have a search that matches 10,000 documents. Then, we can use the categories field as a facet field, so that the result will be (besides the first 10 documents), an overview of the facets, stating that 60,000 of the matching documents have a category from the Crime taxonomy, 80,000 a category from the



Intelligence taxonomy, etc. This is typically useful to enable interactive refinement of the search, whereby a use 'selects' a specific subcategory to make their search more specific.

Since each of the types of extremism considered in CICERO will have different stakeholders and potentially different topics, we have chosen to use different collections of documents for each type of extremism. Similarly, we may decide to have multiple collections to monitor different parts of the CICERO campaign (e.g. one per country, or one per target group).

# IX User Interface

The AW for CICERO web application is available at <u>https://cicero.expertsystemcustomer.com</u>. It requires a username and password to enter and is only available via a secure (encrypted) connection, therefore any content sent between the server and the user is not visible to servers in between.

The first thing a user needs to do is to log into the system:



Users can have different roles:

• **CaseManager/Supervisor**: are responsible for creating an analysis case, i.e. defining which are the questions that need to be answered, what are the sources that could provide the



answer to the question and assign who is responsible for populating the collections, analyzing the content and writing the result for dissemination to decision makers.

- CollectionManager: are responsible for curating the collected content, e.g. removing documents that are not relevant, configuring new sources and loading new documents. In CICERO, they may need to add new sources as new topics become relevant to specific (counter) narratives. E.g. if a shooting like Christchurch occurs, it will be useful to collect content based on keywords related to that event.
- **Analyst**: are responsible for using the search/exploration functionality to gather metrics and write reports that aim to answer the posed analysis questions.
- Customer: these are usually the decision makers. In CICERO they will be those in charge of defining and tweaking the campaign strategy, but also those in charge of monitoring the quality of the campaign. They need to get the results of the analyses made by the analysts in order to make the right decisions based on pre-existing strategies.

Depending on the user's role, they may see different aspects of AW when the log-in.

**Case-managers** and **supervisors** should see a screen summarizing the progress of each of the cases they are managing. E.g.

	Analysts' Workspace	Case + New All Cases >	Q 🖡 Ø (	
::	Semantic Cases Knowledge	Sensors System	CASES PRODUCTION DISSEMINAT	NON
C	Cases	Overall Collections	Case Feeds	
(i)	Filter Cases by Name All Cases	Far right case Colle (486083)	Ecoterrorism 15 days ago 15 days ago	
<b>*</b>	Counter jihadism ① Detail  Peport	Salafi jihadism Coll (33104)	Ronald Denaux (ESI) (rdenaux@cicero.aw.local)	
<ul><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li><li></li></ul>	Ecoterrorism ① Detail  @ Report	Test Jihadism Team C (4394)	Case Team has been edited 15 days ago	
	Far left case            ① Detail          〗 Report	Poirtical Islam Coll (0)	Far right case Fase Team has been edited Anonald Denaux (ESI) (rdenaux@cicero.aw.local)	_
	Far right case ⓓ Detail		Salafi jihadism	
	Incel Extremism      Detail Report	~	Ronald Denaux (ESI) (rdenaux@cicero.aw.local)	~



Among others they are able to monitor the contributions of the team members for each of the cases:



**Collection managers** are able to monitor and configure the sources, webcrawls, and collections as shown in the screenshots below:



Analysts' We	orkspace Sou	urces +	lew All Cases	<b>~</b>							Q		<b>4</b> 0
Sources @Web-Cr	awl 🖉 Collections												
Any time ~ Any Collections ~				Any Type V		Search sources				+ Source			
✓ Far right twitter ✓ Far right twitter ✓ Ø Ø Ø ✓ Salafi jih ✓ Collection: Far right case Collection Collection: Bar right case Collection ✓ Last document update: a few seconds ago ✓ Last document update: a few seconds ago			afi jihadism twi <b>30.</b> ollection: Salafi ast document up	tter V V E 65K jihadism Collection odate: 13 minutes ago	Far right rss 2 2 2 2 6 8.69K Collection: Far right case Collection Last document update: a few seconds ago			BlackPill MGTOW Incel					
Incels.co forums 5.9 Collection: Incel Ex East document up	tremism Collection date: 14 minutes ago	© 1000	smax forum 5.9 Illection: Incel Ex st document upo	Collection Attermism Collection Iate: a few seconds ago	¥ Twitte Coll	ection: Coun document u	d 44K Iter Radicalis pdate: 6 minu	ation utes ago	Collec	i <b>tter user</b> tion: Test J Last docu	s 85 ihadism 1 iment upo	0 Feam Ca late: an	sse Collection
<ul> <li>Far left twitter</li> <li>€</li> <li>€</li> <li>€</li> <li>6033</li> <li>Collection: Far left case Collection</li> <li>▲ Last document update: 4 minutes ago</li> <li>Far left rss</li> <li>€</li> <li>€</li> <li>€</li> <li>€</li> <li>8</li> <li>3001</li> <li>Collection: Far left case Collection</li> <li>▲ Last document update: an hour ago</li> </ul>		<ul> <li>☑ An</li> <li>∴</li> <li>∴</li> </ul>	Animal Liberation Front twit  Animal Liberation Front twit  Animal Liberation Front twit  Collection: Ecoterrorism Collection Collection: Ecoterrorism Collection Collection: Ecoterrorism Collection Last document update: 7 days ago			<ul> <li>▶ test RSS sources</li> <li>▲ ● ▲ ●</li> <li>▲ 4888</li> <li>Collection: Test Jihadism Team Case Collection</li> <li>▲ Last document update: 39 minutes ago</li> <li>▲ Animal rights liberation twit ▲ ● ▲ ●</li> <li>▲ 271</li> <li>Collection: Ecoterrorism Collection</li> <li>▲ Last document update: 2 days ago</li> </ul>			<ul> <li>Salafi jihadism rss</li></ul>				
		■											
From Orgs	22 ter Radicalisation	B 🕀 Fai	right QOF anti- 2 Collection: Far rig	immigrati∕ ⊘ ∕ ₪ 19 ght case Collection		I RightsLib C 2 ection: Ecote	00F a 12 rrorism Colle	ection	€ Fa	Collection	Maxism- 18 Far left o	Leninisr <b>7</b> case Col	n 🖉 🥏 🍠 🔋
Analysts' Wo	orkspace Sou	ırces +	ew All Cases	<b>v</b>							Q		<b>4</b> 0
Sources Web-Cra + Collection	wi Collections Open Export Colle mport Export	ections				<b>D</b> ~	A Num ^	m Crosting	Last Index	Q.			25
Name	Description	🌡 Delegate Adm	nistrator	🌢 Owner	Status	Priority	Docs	Date	Update	Urls	@ Sou	irces	Actions
Counter Radicalisation	Collection of counter radical content	Ronald Denaux (E (rdenaux@cicero	SI) aw.local)	Ronald Denaux (ESI) (rdenaux@cicero.aw.local)	Active	Normal	19516	Jun 27, 2019	10 minutes ago	0	14	+	Actions
Ecoterrorism Collection				Olga Salas (ESI) (osalas@cicero.aw.local)	Active	Low	4767	May 14, 2019	42 minutes ago	0	6	٠	Actions
Far left case Collection		•		Olga Salas (ESI) (osalas@cicero.aw.local)	Active	Low	1690	May 14, 2019	7 minutes ago	0	5	٠	Actions
Far right case Collection				Olga Salas (ESI) (osalas@cicero.aw.local)	Active	Low	486149	May 14, 2019	4 minutes ago	0	8	٠	Actions
Incel Extremism Collection				Ronald Denaux (ESI) (rdenaux@cicero.aw.local)	Active	Low	67095	Jun 27, 2019	3 minutes ago	0	11	+	Actions
Political islam Collection				Olga Salas (ESI) (osalas@cicero.aw.local)	Active	Low	-	May 14, 2019		0	0	+	Actions
Salafi jihadism Collection				Olga Salas (ESI) (osalas@cicero.aw.local)	Active	Low	33104	May 14, 2019	16 minutes ago	0	4	+	Actions
Test Jihadism Team Case Collection		-		Ronald Denaux (ESI) (rdenaux@cicero.aw.local)	Active	Low	4394	May 08, 2019	42 minutes ago	0	4	÷	Actions

Analysts can explore and search the collected documents, they can use a variety of filters based on the semantic analysis to make it easier to narrow down their document exploration to just those that are relevant for answering a particular question as part of a case. E.g. they may want to only look at

Previous 1 Next

Showing 1 to 8 of 8 entries



documents in the counter-radicalisation collection, which have predominantly a negative sentiment, mention, discuss a moral obligation (from the strategic radicalization narrative taxonomy) and mention a place in Belgium.



The AW interface provides a variety of diagrams and graphs for summarizing the content of the selected documents, for example, based on a search with 4,754 documents, it is able to generate a summary of the sentiments and emotions:



Analysts' Workspace Explore Q 1 🔺 🛛 🕘 + New All Ca Topics Entities Emotions Relations Tags Places Social Media Claim Alerts Kno ledge 🕕 🛛 Notes <mark>8</mark>2 D Filters T Filter by Explore Sources 🗻 🗸 Any File Type ~ Sorted by Publis... Emotions Show All Any time 🗸 Counter Radicali.. Any KBs 🗸 Any Type 🗸 173 Love Q 4,754 results for all contents 140 Sadness Action 117 Success 105 86 Anger polit. Desire Action Sentiments Sadness Hatred Anger onsiderati Relax neutral 358 Spirituality Surprise negative 305 Success Passion C positive 287 Commitment Pride Sympa... C Shame Anxiety Suffering Torment Hope Joy Love Remorse Offe Repulsion Affection Anguish Exas Document Charts Photo Archives - Page 5 of 5 - We have the choice

Similarly, it can generate tag-clouds allowing to see what other words are relevant for the selected subset of documents:

opics Entities Emotions Re	lations Tags	Places Social Media Claims Alerts Knowledge 💽 Notes 6 2 2 1 0 Saved Search
Filters Tilter by		Explore
Main Elements	Show All	Any time v Counter Radicali Any KBs v Sources 🕦 v Any Type v Any File Type v Sorted by Publis
choice	4,204	Q 4,754 results for all contents
have the choice OR NOT	3,043	
circle	2,365	initiany and
photo	1,252	
johan van steen	1,160	لله المعالم الم معالم المعالم ال
		And the property of the proper



#### Another useful diagram is that of the places mentioned in the selected content:





Analysts have access to their own personal notebook to copy diagrams and make notes of interesting documents and topics. Eventually, they can copy content from their notebooks into a report that summarises their findings and can be used to answer the case questions:



Finally, the **customers** receive the created reports and can provide feedback which may trigger subsequent analyses.



# 3. BASELINE EVALUATION

# 3.1 Monitored Sources and Existing Campaigns

As of July 25<sup>th</sup> 2019, we have gathered 19,518 documents related to pre-Cicero counter-radicalisation campaigns. 18,046 of those have been automatically analysed by Cogito.

#### Sources

For Twitter:

- Twitter counter rad: tweets mentioning accounts and hashtags:
  - #deradicalization, @WwB\_SAVE, @GIRD\_S,
     #WomenWithoutBorders, @ISDglobal
- From Orgs: tweets from accounts
  - WwB\_SAVE, GIRD\_S, ISDglobal, info\_radical or preventviolence
- To Orgs: tweets directed at or replying to accounts:
  - WwB\_SAVE, GIRD\_S, ISDglobal, info\_radical or preventviolence
- #EtSiJavaisTort: tweets mentioning the hashtag for the <a href="http://etsijavaistort.org/">http://etsijavaistort.org/</a> campaign
- #WhatIfIWasWrong: English version for the <u>http://etsijavaistort.org/</u> campaign

RSS feeds:

- Efus fr feed: French documents and comments for <a href="https://efus.eu/fr">https://efus.eu/fr</a>
- Efus en feed: English documents and comments for <u>https://efus.eu/en</u>
- Women without borders: either from <u>http://womenwithoutborders-save.blogspot.com</u> or <u>https://www.women-without-borders.org/</u>
- etsijavaistort feed, documents from <a href="http://etsijavaistort.org">http://etsijavaistort.org</a>
- ISD.org feed: document from <a href="https://www.isdglobal.org">https://www.isdglobal.org</a>

A webcrawl for pages:

- <u>http://etsijavaistort.org/</u>
- <u>http://wehavethechoice.com/</u>
- <u>http://www.garance.be/</u>
- <u>https://efus.eu/en/</u>

source	docs
"webcrawl"	16579
"twitter"	1409
"rss"	152
"web"	135
"image"	74



- <u>https://savebelgium.org/</u>
- <u>https://violence-prevention-network.de/</u>
- <u>https://www.hherald.org/</u>
- <u>https://www.isdglobal.org/</u>
- <u>https://www.keuzekompas.be/</u>
- <u>https://www.operation250.org/</u>
- <u>https://www.pluralismoyconvivencia.es/</u>
- <u>https://www.thefuturegeneration.nu/</u>

The 18K collected documents come mostly from the webcrawl, followed by Twitter and the rest of the sources. In terms of sites, efus.eu and wehavethechoice.com are the largest sources, followed by garance.be and Twitter, as can be seen in the following table.

# 3.2 Social media analytics

The online campaign carried out by CICERO will mainly operate on the most common social media (i.e. Facebook and Twitter), since they provide a significantly more consistent audience and can potentially guarantee a much broader reach.

Furthermore, both these social media analytics tools are available and openly accessible by anyone.

source_id	docs
"etsijavaistort.org"	780
"#EtSiJavaisTort"	3
"Women without borders"	25
"wehavethechoice.com"	4326
"hherald.org"	4
"thefuturegeneration.nu"	27
"Twitter counter rad"	1341
"efus fr feed"	45
"efus en feed"	56
"etsijavaistort feed"	14
"pluralismoyconvivencia.es"	401
"#WhatIfIWasWrong"	3
"garance.be"	1382
"keuzekompas.be"	112
"efus.eu"	7800
"violence-prevention-network.de"	769
"ISD.org feed"	12
"From Orgs"	199
"operation250.org"	58
"To Orgs"	72



# / Facebook

For Facebook, CICERO will employ Facebook Audience Insights<sup>8</sup>, a tool that is able to highlight all the major information about the audience reached, such as demographics (including age and gender), likes, location and kind of activity of the viewers.

#### // Twitter

The CICERO campaign on Twitter will be evaluated through the use of Twitter Analytics<sup>9</sup>. This tool is able to assess how a Twitter account has performed, especially in terms of profile visits, growth of the followers' base, tweet impressions, and mentions. Additionally, it provides also engagement insights, such as the performance in terms of re-tweets.

# 3.2 Text analytics

# / Socio-demographic analysis

The languages, education levels and estimated age for the sources described in 3.1 are depicted here.

We see that we have a mixture of content in English (60%), French(20%), Dutch(10%) and Spanish (5%). It may be good to get more sources of counter-narrative campaigns in Italian as this will allow us to compare content related to pre-existing campaigns with content related to CICERO's campaign.

Being able to detect the percentage of content in specific languages will enable us to measure whether targets in the various languages are being met.



<sup>8</sup> <u>https://www.facebook.com/business/news/audience-insights</u> [Accessed July 30, 2019]

<sup>&</sup>lt;sup>9</sup> <u>https://analytics.twitter.com/about</u> [Accessed July 30, 2019]





Regarding education level, we see that such campaigns tend to use simple language with over 50% of the overall content at a primary level, and only about 5% at a university level. Some notable exceptions, which tend to use more complex language are operation250.org and women without borders. Examples of sources that use very simple language are ISD.org and wehavethechoice. Right in the middle and clearly geared towards teens is thefuturegeneration.nu which predominantly uses language at the middle and high-school levels, but also efus tends to do this. Assessing the education level of the content produced by a campaign can be a useful way to measure whether the content produced is in line with and appropriate for the defined target audience of the campaign.



Regarding age groups, we see that counter-radical content is written by about 20% of young people and about 75 of middle aged people. Some of the sources were young people provide more content are efus.eu (almost 30%), etsijavaistort.org and keuzekompas.be.

Being able to give an estimate of the age-group of the authors of content will be useful for measuring whether the campaign content is being written in a style that appeals to the target groups. For example, if we decide to target young people in the campaign, we could use this tool to estimate whether sufficient young people are contributing to the CICERO campaign, or whether the content is only being produced by older people and thus may be less appealing to target age-group.



#### // Topic and Sentiment analysis

We can use Cogito's semantic analysis to analyse whether certain content, which is part of a campaign, is staying on-topic and is using the right combination of sentiments. For example, by using the Terrorism taxonomy described in 2.2.IV, we can generate the following graph:



#### % of docs with Terrorism topics per source



The graph shows in the y axis the percentage of documents in a specific source that talks about the topics mentioned at the top of the graph. We see for example that about 16% of the documents in the violence-prevention-network.de source talk about *Religiously inspired terrorism*. Likewise, we see that about 10% of the documents in garance.be talk about *Terrorist attacks on civilians*. Depending on the strategies of each campaign, this may or may not be the desired percentage.

The Cogito results of the sentiment analysis are likewise bound to be useful for determining whether the content being produced and shared on-line are adhering to the targets specified in the campaign strategy. For example, on the collected pre-CICERO counter-radicalisation campaigns we see the following results:



The graph shows at a glance that around 16% of the documents in the garance.be collection contain language related to hatred, but also about 10% related to success. The violence-prevention-network.de shows a similar profile, while other campaigns use less emotions, or focus more on positive emotions like love and satisfaction. Further graphs for other emotions confirm that garance.be uses a much more emotional approach than the other campaigns being depicted.



# % of docs with Emotions topics per source



# % of docs with Emotions topics per source





# % of docs with Emotions topics per source



The graphs above show a good for comparing different sources, but as shown in 2.2.VII the AW is also able to generate an overview of all the sentiment and emotions within a particular cross-section of a collection. Both of these visualisations can provide valuable insights into whether the content produced as part of a campaign (and the resulting responses on social media) are striking the right topics and sentiments.



### 4. CONCLUSIONS AND FUTURE WORK

In this first of three reports on Campaign Effectiveness Evaluation, we focused on describing some of the tools that we have been configuring and testing to be prepared to evaluate the CICERO campaign once it gets started. Since some details about the campaign's content and dissemination strategy are still being decided, it was not possible to apply the tools to directly relevant content. However, we have managed to collect content about pre-CICERO counter-radicalisation campaigns and have applied the tools to validate that they are able to produce valuable quantitative and qualitative metrics about such campaigns.

As the CICERO campaigns content and dissemination strategy details crystallize, we will configure these and further tools to be able to start collecting data as soon as decisions are being made. That will ensure that, at least for the most automated monitoring part, we are able to successfully monitor and provide feedback about the campaigns' impact and effectiveness. Naturally, besides the tools presented here, we will follow the evaluation methodology as presented in D5.1 to also collect off-line feedback about the campaigns' results, which will also be reported in the second and third Campaign Effectiveness Evaluation Reports.

# CICERO - Counternarrative Campaign for Preventing Radicalization



This project was funded by the European Union's Internal Security Fund —Police

under GA N° 812613

The Commission is not responsible for any use that may be made of the information contained therein, the sole responsibility lies with the authors.